Two satires on failure of 16S rRNA sequences as "Rosetta Stone" records of the evolutionary relationships of bacterial species

by Howard Gest
Distinguished Professor Emeritus
of Microbiology
Adjunct Professor, History and
Philosophy of Science
Indiana University (Bloomington)
gest@indiana.edu

Report from the year 2025 meeting of the American Microbiological Society: *Discovery of the bacterial 'taxonomy gene'*

Originally published in MICROBIOLOGY TODAY, the magazine of the Society for General Microbiology, vol. 27, February 2000, pp. 28-30

Gest's Postulates

Originally published in AMERICAN SOCIETY FOR MICROBIOLOGY NEWS, vol. 65, p. 123, 1999

Report from the year 2025 meeting of the American Microbiological Society: *Discovery of the bacterial 'taxonomy gene'*

Howard Gest

The following satire is based on Howard Gest's view that the evolutionary history of bacteria was more complex than commonly supposed and cannot be traced with accuracy using 16S rRNA sequences as the sole criterion. Indeed, during recent years, a number of reports summarizing new research findings, including evidence suggesting the widespread occurrence of lateral gene transfer, cast doubt on the validity of bacterial evolutionary phylogeny based on rRNA trees. Gest emphasizes that rRNA sequences will probably prove to be useful eventually for identifying certain kinds of taxonomic relationships, but will not serve to provide an unambiguous evolutionary phylogeny of bacteria. Accordingly, he argues that changing the names of numerous well-known genera and species on the basis of rRNA sequences is premature and counter-productive to formulation of a logical, scientific scheme of bacterial relationships and classification.

It is not widely known that while Sir Isaac Newton was developing his epoch-making work in mathematics and astronomy, he spent much effort and time on alchemical, esoteric, theological and mystical studies. John M. Keynes estimated that Newton left behind unpublished manuscripts of more than 1,000,000 words in a great box at Cambridge. According to Keynes, Newton looked on the whole universe '*as a riddle, as a secret which could be read by applying pure thought to certain evidence, certain mystical clues*'. Newton believed that these clues were '*partly in certain papers and traditions handed down by the brethren in an unbroken chain back to the original cryptic revelation in Babylonia*'. There is no doubt that through the ages, lesser minds than Newton's made similar unsuccessful attempts to unlock the secrets of life and the universe from ancient writings.

Etruscan promises

The Etruscan language has defied all attempts at translation. During the 1990s, a bronze sheet with Etruscan inscriptions dating from the late third to the early second century BC was discovered near Cortona (Italy) and this gave some promise for progress (*Italy Daily*, July 1, 1999). However, the limited text of only 27 previously unknown words proved to be a disappointment because it appeared to concern only a transaction of some sort between a few aristocratic families. Hopes soared in 2008 when a more detailed tablet was unearthed beneath the Basilica San Vitale near Ravenna. The 'San Vitale Slab' was inscribed in presumably parallel messages in Greek, Latin and Etruscan, and is now in the British Museum, next to the Rosetta Stone. Only part of the San Vitale Slab could be deciphered and surprisingly, this dealt with scientific matters. The Etruscan script gave the relative atomic weights of selenium, platinum, plutonium and seaborgium (element 106), and the value of π to 987 places. Of course, the excitement in academic circles was considerable, but further progress again slowed to a snail's pace. Nevertheless, the advance was a harbinger showing Newton's prescience in seeking explanations of natural phenomena in ancient records.

Concealed messages in DNA

The possibility that universal truths could have been deliberately concealed in 'DNA language' was bolstered on the eve of the present millennium by scientists who devised an encryption system based on genetic coding. The method developed by Clelland *et al.* utilized features of a procedure used by German spies during World War II to transmit secret information. In the spy system, the photograph of a typewritten page was greatly reduced to the size of a 'microdot' that could be pasted over an ordinary dot (i.e. a full stop) in a seemingly innocuous letter. The concealed message could easily be recovered by photographic enlargement.

Clelland *et al.* went many steps further by using a DNA-encoded message that could be camouflaged within the enormous complexity of human genomic DNA and then further concealed by confining the message within a microdot. The encryption key was simple. Letters of the English alphabet were assigned to nucleic acid base triplets. Thus letter *a* = CGA, *b* = CCA, *c* = GTT, etc. Primers were designed so that the base-encoded message in a polynucleotide could be selectively revealed using the PCR reaction directly on the microdots. The Clelland method was successfully demonstrated using 'DNA microdots' pasted on full stops (with common adhesives) in a printed letter sent through the US mail.

The DNA microdot method seemed to be foolproof, but puzzles based only on technological gimmicks obviously can be expected to be solved using even more advanced technological tricks. Thus in 2017, Valentin Miescher and his colleagues in Switzerland finally

discovered the functions of so-called junk DNA (and thereby the so-called Second Order Genetic Code) and this, together with ultra-sensitive scanning methods, quickly led to simple ways of locating and decoding the DNA microdots. It then became evident that construction of more sophisticated coding systems would require input of esoteric cultural information as well as advanced micro-technological innovations.

Maturation of the Glass Bead Game

After a lapse of some 250 years we now have evidence that Newton was on the right track in searching for explanations of natural phenomena in very ancient sources. Some success has finally been achieved as a consequence of development of the Glass Bead Game (GBG) invented in 1943 by Hermann Hesse (Nobel Laureate 1946).

The elaborate GBG was originally practised by a select monastic-like brotherhood. The game required high intelligence and great skill, and one could become a 'Magister' only after many years of training and trials. An early, primitive version of the GBG was in the form of musical exercises played on a frame holding several dozen wires on which could be strung glass beads of various sizes, shapes and colours. The wires corresponded to musical staffs and the beadsto time-values of the notes. With this device, a person could *'represent with beads musical quotations or invented themes, could alter, transpose, and develop them, change them and set them in counterpoint to one another'*. Within a few decades, the GBG was taken over by mathematicians, who modified it to a high degree of flexibility. Soon, the GBG was developed so that *'it was capable of expressing mathematical processes by special symbols and abbreviations. The players, mutually elaborating these processes, threw these abstract formulas at one another, displaying the sequences and possibilities of their science.'*

It was only a matter of time before the GBG was taken up and imitated by other scientific and scholarly disciplines. Eventually, the GBG could be regarded as a kind of language. According to the World Commission of the GBG, its archives contain the *'register of all hitherto examined and accepted symbols and decipherments, whose number long ago exceeded the number of the Chinese ideographs'*. With the discovery of the basic structure of DNA by Watson & Crick in 1953, it could be expected that 'DNA language', and thus 'protein language', would furnish a lexicon for variations of the GBG.

The first 'protein-language GBG' emerged in the 1990s. To compare patterns of amino acid sequences of various proteins, individual English letters had earlier been assigned to each of the ca 20 amino acids. Inevitably, someone posed the questions: do any English words occur in such sequences, and if there are English words, what is the longest one can find? Game players in certain organizations had the spare time, necessary computers and funds to play the game, informally known as 'protein talk'. In 1993, two Swiss scientists (Gonnett & Benner) matched the entire *Oxford Unabridged English Dictionary* (2nd edition; 20 volumes; 572,728,830 characters, with information content close to that of the human genome) against the entire SwissProt protein sequence database. They found two words with nine characters: *hidalgism* (the manner or practice of a hidalgo, a man of the lower nobility in Spain) and *ensilists* (plural of ensilist, one who protects his crops by ensilage). The game players concluded, *'In addition to being the longest strings appearing simultaneously in the English and protein languages, these are also candidates for the most unusable pieces of information simultaneously in lexicography and in biochemistry'*.

The protein talk report cited soon elicited a response from a British scientist (David Jones) who asked the sensible question, *'But what of other languages? Given that the ownership of the longest peptide-word* [i.e. protein-word] *will undoubtedly become a source of intense national pride, I thought it wise to investigate.'* So, Jones proceeded to search the SwissProt database with a multilingual word list of 1·3 million words from Danish, Dutch, English, Finnish, French, German, Italian, Norwegian, Swedish, Spanish and some Esperanto. Results: the nine-letter words: *ansvarlig* (Danish for 'liable'), *haletante* (French for 'breathless'), *salasivat* (the past tense of Finnish for 'to keep hidden' or 'to encode'), *saltsilda* (Norwegian for 'salted herring') and *stillassi* (the perfect subjunctive of the Italian word *stillare*, 'to drip'). The search also turned up one 10-letter Italian word, *annidavate* (past imperfect tense of *annidare*, 'to nest'). Jones asked, *'How long will we have to wait before Germany finally scoops the honours with the possible 27-letter peptide word for social sciences: Gesellschaftswissenschaften?'* Subsequent research (2000–2010) on large genes and proteins indicated that their sequences might well encode even longer messages intelligible in human language scripts.

LEFT:
The Rosetta Stone
PHOTO COURTESY THE BRITISH MUSEUM

By 2015, the GBG had a special language and set of rules for every discipline and subdiscipline, and so a wide variety of derivative GBGs evolved. The further development of computers greatly enhanced biological variations of the GBG. But a major problem in solving puzzles persisted, namely identification of the human language in which ancient sages could have encoded particular 'Secrets of Life'. Even with supercomputers, it would require vast knowledge of human history and culture and decades of trials to identify the languages in which explanations of biological phenomena were hidden in ancient writings.

Breakthrough in 2025

Fortunately, serendipity continues to play a significant rôle in scientific discovery. One member of a large team of scientific workers at NASA's PABLUM project (Planetary AstroBiological Launch for Unidentified Microbes) happens to be married to a Chinese history scholar who is familiar with the history of science and civilization in China (see Joseph Needham's monumental compilation). At her suggestion, the team, applying principles of the GBG, checked a rare Chinese dialect used by followers of Mencius (ca 371–289 BC), the so-called Second Sage, who was second only to Confucius, the Supreme Sage. As announced in the Abstracts of this year's meeting of the American Microbiological Society, this led to identification of the hitherto unknown 'taxonomy gene'.

When the base sequence of the taxonomy gene of a bacterium is transliterated into the Mencius dialect, and then transmogrified into English equivalents, the true name of the organism emerges. Thus, its evolutionary phylogeny is revealed and century-long controversies on the most relevant criteria for constructing taxonomic schemes are at an end. From now on, *Escherichia coli* will be designated *Proteofermentoformicus lipocylindricus*. Similarly, a majority of existing genera and species names will be changed by a reconstituted taxonomy committee responsible for the new *Bergey's Ultimate Manual of Definitive Bacteriology*.

How did the taxonomy gene originate?

Francis Crick's book *Life Itself* gives food for thought. Crick entertains the possibility that on some distant planet there '*evolved a form of higher creature who, like ourselves, had discovered science and technology, developing them far beyond anything we have accomplished...*'. Imagining various disaster scenarios, Crick suggests that these creatures would have planned to colonize other planets, and that they may have sent microbes to the Earth as an initial step. With his classic style of understatement, Crick says, '*The senders could well have developed wholly new strains of micro-organisms, specially designed to cope with prebiotic conditions, but whether it would have been better to try to combine all the desirable properties within one single type of organism or to send many different organisms is not completely clear*'.

It would not be surprising that a 'supercivilization' of the kind envisioned by Crick could have deliberately designed genes (for implantation into microbes) containing messages of importance for edification of the higher forms of life that would eventually evolve on Earth. Discovery of the bacterial taxonomy gene encoded in an obscure Chinese dialect lends further support to the increasingly popular view that representatives of very advanced civilizations visited Earth eons ago, and more recently from time to time.

Further reading

Clelland, C.T., Risca, V. & Bancroft, C. (1999). Hiding messages in DNA microdots. *Nature* 399, 533–534.

Crick, F. (1981). *Life Itself – Its Origin and Nature*. New York: Simon & Schustet.

Gest, H. (1999). Bacterial classification and taxonomy: a 'primer' for the new millennium. *Microbiology Today* 26, 70–72.

Gonnet, G.H. & Benner, S.A. (1993). A word in your protein. *Nature* 361, 121.

Hesse, H. (1970). *Magister Ludi – The Glass Bead Game* (English translation). New York, Toronto, London: Bantam Books.

Jones, D. (1993). More protein talk. *Nature* 361, 694.

Keynes, J.M. (1956). In *Essays and Sketches in Biography*, pp. 280–290. New York: Meridian Books.

Needham, J. (1954–1988). *Science and Civilisation in China*. Cambridge: Cambridge University Press.

Letters

Gest's Postulates

My letter is in the form of a satire which addresses serious issues in the debate as to (a) whether 16S rRNA sequences found in natural sources actually signify extant bacterial species, (b) the legitimacy of using only 16S rRNA sequences as the basis for changing the names of bacterial genera and species, and (c) the taxonomy of bacteria. This is a memorandum issuing from the recently established Commission on Taxonomic Justice. The Commission deals with instances of taxon piracy (frivolous changes of genus and/or species names) and other Infractions of the Nomenclature Code (10th revision).

Gest's Postulates relate to free-living bacteria, and refer to "virtual bacteria" reported to exist in natural sources as indicated by detection of 16S rRNA sequences ("virtuals" are sometimes referred to by the cognoscenti as "computer bacteroids").

The Postulates are designed to forestall premature corruption of *Bergey's Manuals* (of Actual Bacteria): First, to register a 16S sequence with a generic or species name, the author must provide evidence that the sequence is (a) not a chimeric molecule due to PCR coamplification of 16S rRNA genes from mixed bacterial genomes (see Wang and Wang, Appl. Environ. Microbiol. **63**:4645, 1997), (b) not due to common bacterial contaminants in reagents used to prepare genomic DNA (Tanner et al., Appl. Environ. Microbiol. **64**:3110, 1998), (c) not due to human commensals from saliva, exhalations, perspiration, or fingerprints, and (d) not due to laboratory variants of well-characterized species.

Second, designation of a "virtual" as "uncultivable" on the basis of observing no growth upon addition of yeast extract or a similar complex supplement to a so-called standard medium will not be accepted by CANCELD (Committee Advising Nomenclature Corrections Eliminating Literature Deficiencies).

Third, applying the "3/5" rule, failure to provide evidence for existence of existence of a corresponding cellular entity within three years will in each instance oblige the author to publicly rescind the last five "16S virtuals" that he/she has designated in publications. Three such episodes will be declared a taxonomic felony, and after a grace period of one year will result in prohibition from obtaining cell cultures from the ATCC and cooperating culture collections.

A taxonomic felony can be erased only if the author satisfactorily passes an examination (true/false) administered by the Judge Advocates Office of Nomenclature Reform. The examination is based on classics of microbiology, including: B.C.J.G Knight, *Bacterial Nutrition* (His Majesty's Stationery Office, 1938); M.W. Beijerinck, Verzamelde Geschriften van M.W. Beijerinck, ter gelegenheid van zijn 70sten verjaardag met medewerking der Nederlandsche regeering uitgegeven door zijne vrieden en vereerders, 6 volumes (Martinus Nijhoff, 'S Gravenhage, 1921–1940); and S. Winogradsky, Microbiologie Du Sol (Masson et Cie, Paris, 1949). Erasure of a taxonomic felony also requires authors to state in succeeding papers dealing with "virtuals" the following observation of Noberto Palleroni (*ASM News*, October 1994, p. 537): "No shortcuts have ever been invented to bypass the integral study of the whole cell, since diversity is, of course, something more than a set of variations on the theme of a macromolecule."

Authors are warned that CANCELD takes a dim view of gratuitous comments of the kind made by Palleroni and Doudoroff in 1965 describing the "Identity of *Pseudomonas saccharophila*" (J. Bacteriol. **89**:264). Palleroni and Doudoroff correctly pointed out that "Unfortunately, in spite of the fact that more than 50 reports dealing with this monotypic species have been published in the course of the past 24 years, neither its name or its description (Doudoroff, Enzymologia **9**:59, 1940) appears in *Bergey's Manual of Determinative Bacteriology*." Following a detailed description of the bacterium, however, they added: "We also offer the following information for what it may be worth for taxonomic keys: not known to be pathogenic for whales or sweet potatoes; production of indole on egg nog agar not reported."

Howard Gest
Indiana University, Bloomington